Office of
PUBLIC SECTOR INFORMATION

reflections on civic information

john sheridan
head of e-services
the office of public sector information

john.sheridan@opsi.x.gsi.gov.uk

Office of
PUBLIC SECTOR INFORMATION

the office of public sector information

- At the heart of information policy with a public sector wide remit
- Responsible for the management of Crown copyright
- Responsible for the superintendence of a range of official publishing activities, including all UK primary and secondary legislation and the official Gazettes
- Provides online access to legislation – delivering one of government's most used online services
- Encourages the re-use of public sector information

Office of
PUBLIC SECTOR INFORMATION

public sector information **is vital**

- Organisations in the public sector are key suppliers of reliable base information, on which both a knowledge economy and an information society are built
- PSI underpins numerous information value chains
- This gives us both tremendous leverage and a great responsibility
- Technological innovation is driving the re-use agenda forward

Office of
PUBLIC SECTOR INFORMATION

a growing **debate**

- There is a growing public debate about the role, importance and potential of PSI. This is a good thing!
- The Guardian “Free Our Data” campaign is increasing awareness, but is it a symptom or a cause?
- This debate is beginning to reach a wider audience that needs to be informed
- The Office of Fair Trading has conducted a market study into Public Sector Information. Government response is due for publication late Spring / early Summer 2007.

Office of
PUBLIC SECTOR INFORMATION

the web changes **everything**

Office of
PUBLIC SECTOR INFORMATION

that's **everything**

- Fighting wars
- Winning elections
- Making scientific discoveries
- Buying and selling (anything)
- Making money
- Working
- Entertaining ourselves
- Falling in love
- …

Office of
PUBLIC SECTOR INFORMATION

audacious, perhaps **inconceivable**

"*We've produced an information infrastructure that few would have anticipated, with the possible exception of Vannevar Bush, but I think even he would have thought the scale of all this extraordinary. Fifty years ago, it might have appeared audacious, perhaps even inconceivable, that we could have built the kind of global infrastructure that now surrounds us*."

Prof Nigel Shadbolt

WWW2006, Edinburgh, May 2006.

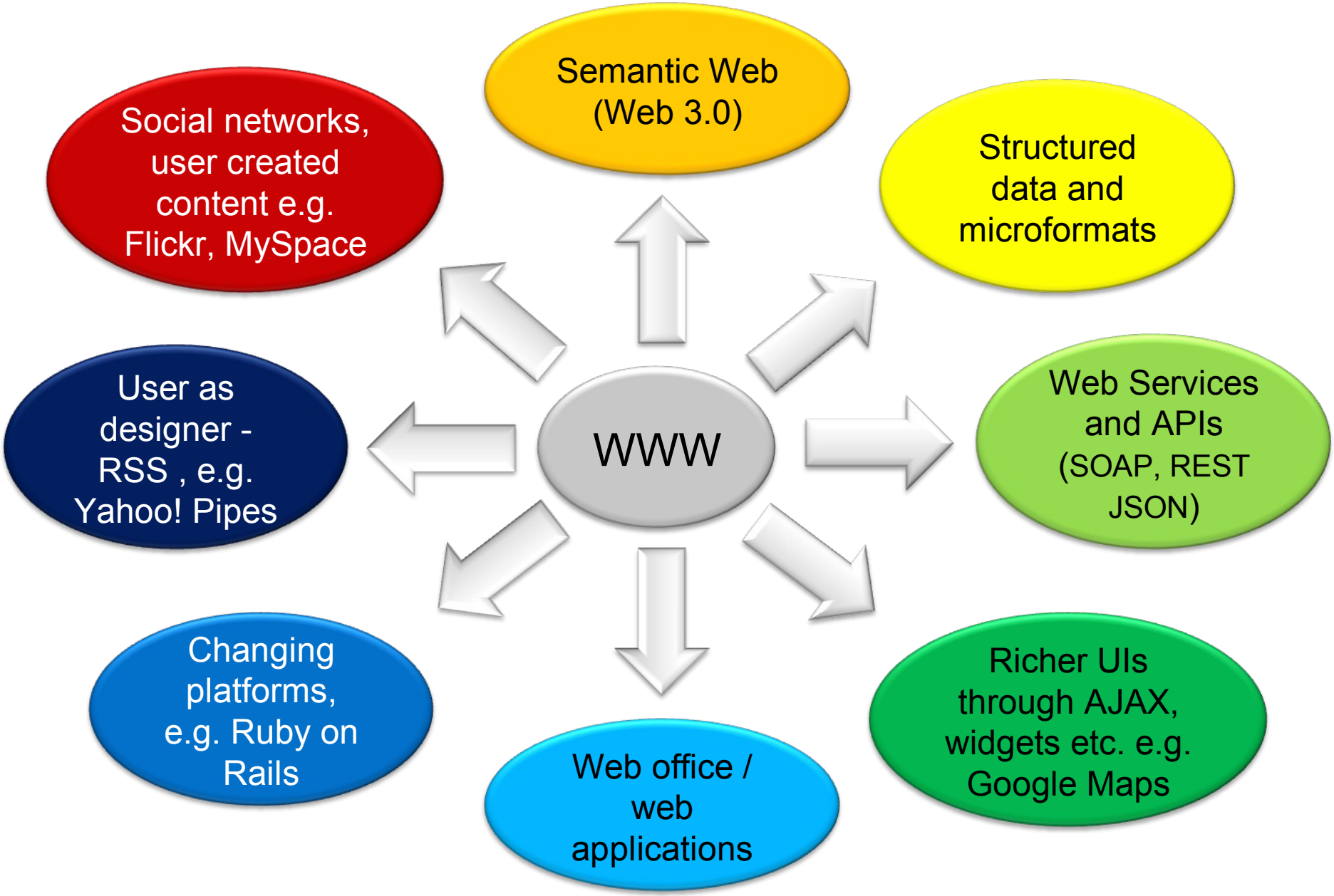
Office of
PUBLIC SECTOR INFORMATION

all from something **so simple**

"The universality and flexibility of the Web's linking architecture has a unique capacity to break down boundaries of distance, language, and domains of knowledge."

Tim Berners-Lee

United States House of Representatives Committee on Energy and Commerce Subcommittee on Telecommunications and the Internet
Hearing on the "Digital Future of the United States: Part I -- The Future of the World Wide Web", March 2007


Semantic Web (Web 3.0)
Social networks, user created content e.g. Flickr, MySpace
Structured data and microformats
User as designer - RSS , e.g. Yahoo! Pipes
WWW
Web Services and APIs (SOAP, REST JSON)
Changing platforms, e.g. Ruby on Rails
Web office / web applications
Richer UIs through AJAX, widgets etc. e.g. Google Maps

Office of
PUBLIC SECTOR INFORMATION

towards a web **of data**

- *Today, the Web is quite effective at helping us to publish and discover documents, but the individual information elements within those documents (whether it be the date of any event, the price of a item on a catalog page, or a mathematical formula) cannot be handled directly as data. Today you can see the data with your browser, but can't get other computer programs to manipulate or analyze it without going through a lot of manual effort yourself. As this problem is solved, we can expect that Web as a whole to look more like a large database or spreadsheet, rather than just a set of linked documents.*

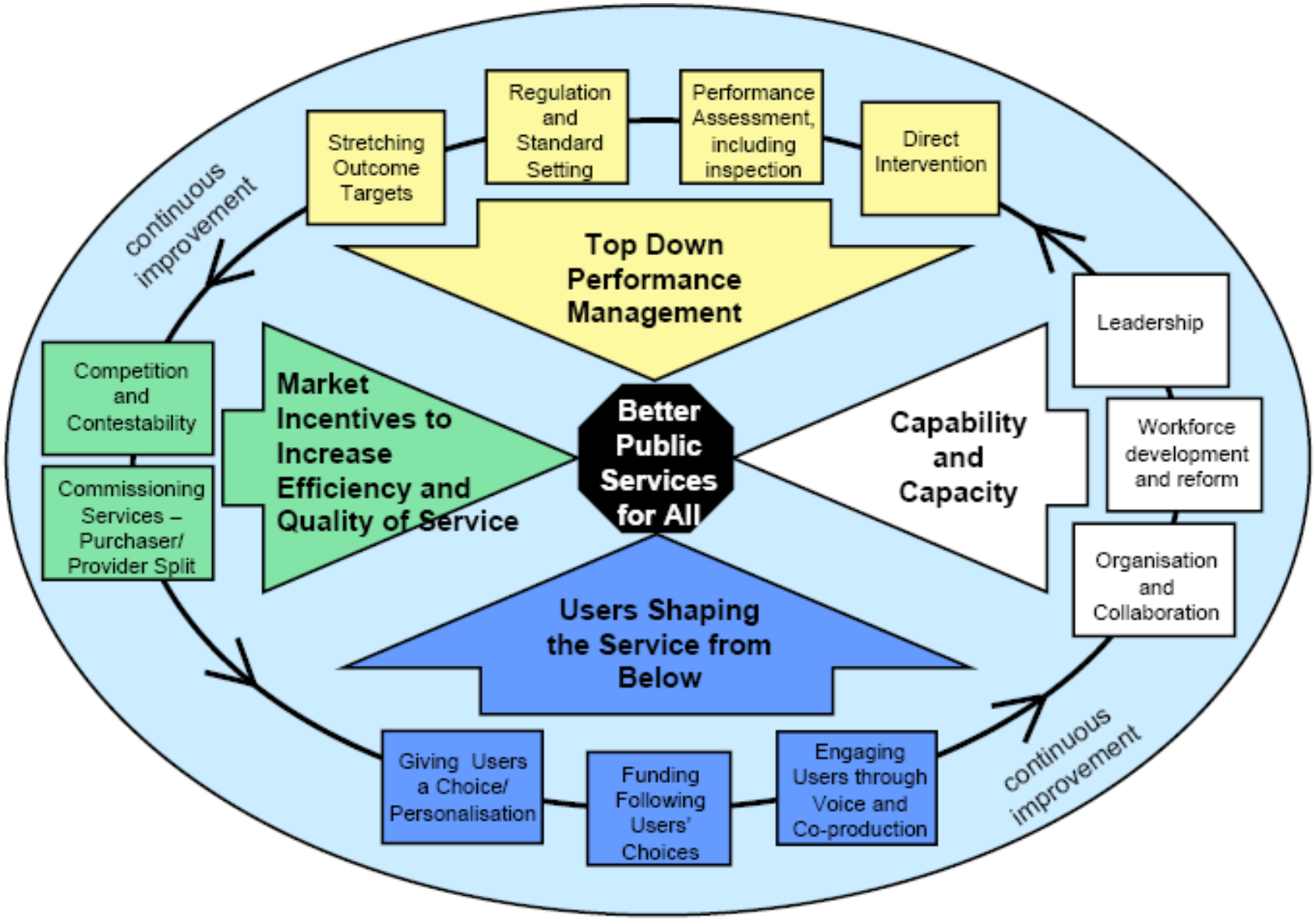
Tim Berners-Lee

United States House of Representatives Committee on Energy and Commerce Subcommittee on Telecommunications and the Internet Hearing on the "Digital Future of the United States: Part I -- The Future of the World Wide Web“, March 2007.

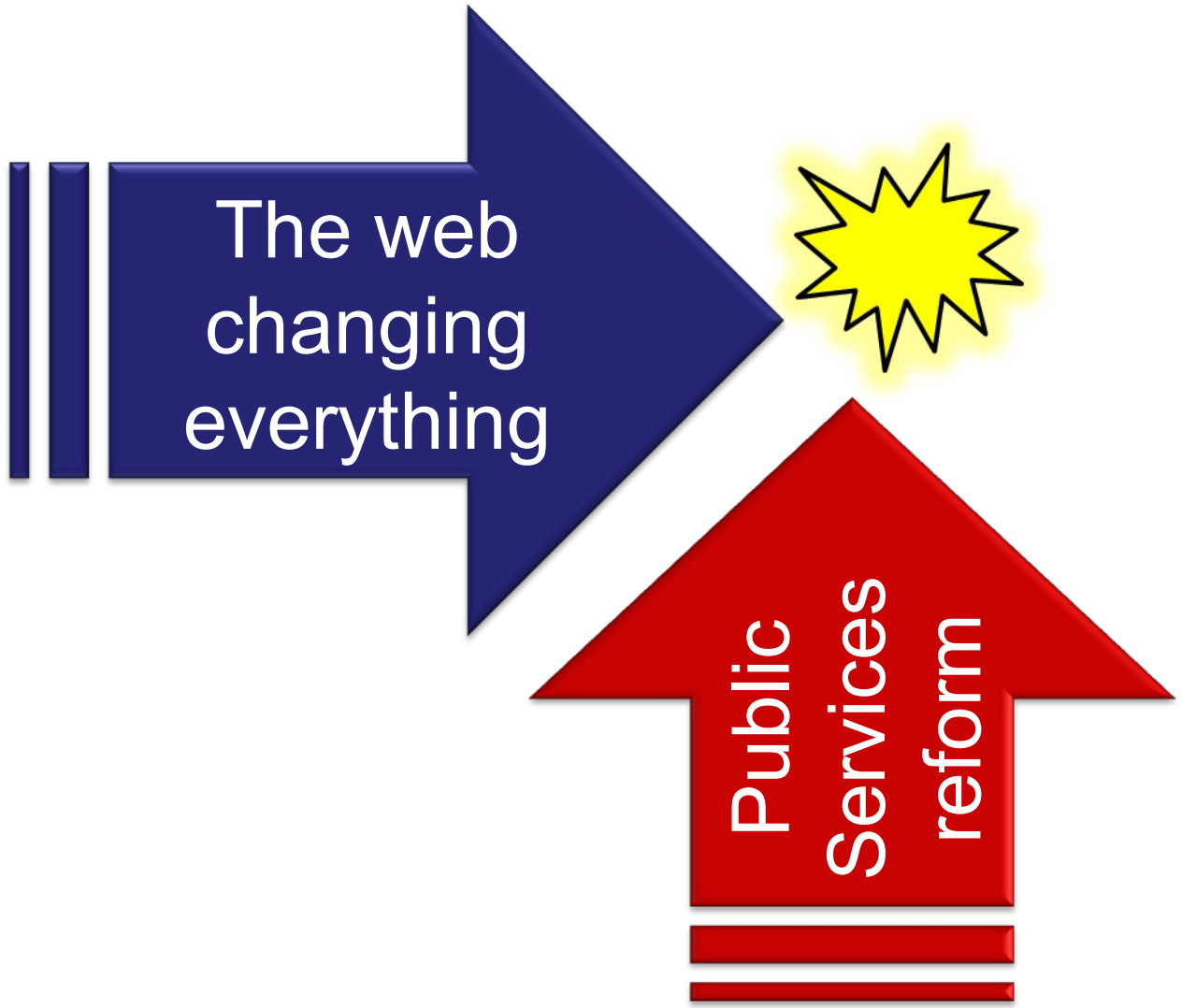
Office of
PUBLIC SECTOR INFORMATION

meanwhile…

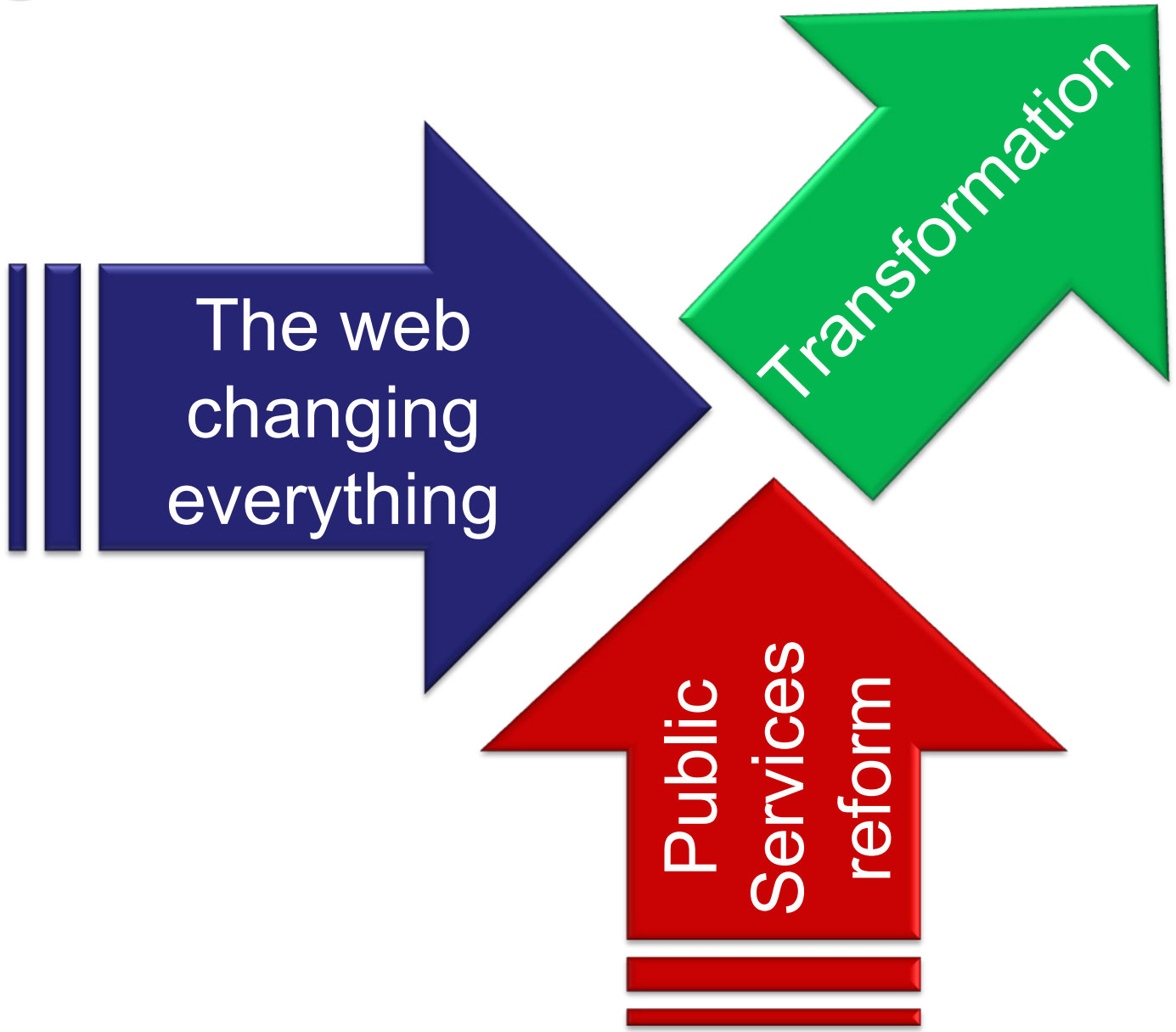
Government's model for public service reform


Stretching Outcome Targets
Regulation and Standard Setting
Performance Assessment, including inspection
Direct Intervention
continuous improvement
Top Down Performance Management
Leadership
Competition and Contestability
Market Incentives to Increase Efficiency and Quality of Service
Better Public Services for All
Capability and Capacity
Workforce development and reform
Commissioning Services – Purchaser/ Provider Split
Organisation and Collaboration
Users Shaping the Service from Below
Giving Users a Choice/ Personalisation
Funding Following Users' Choices
Engaging Users through Voice and Co-production
continuous improvement

Office of
PUBLIC SECTOR INFORMATION



Office of
PUBLIC SECTOR INFORMATION



Office of
PUBLIC SECTOR INFORMATION

provide and enable

Office of
PUBLIC SECTOR INFORMATION

re-use and the **semantic web**

- *Definition*: The **Semantic Web** is the representation of **data** on the World Wide Web.
- The semantic web is about allowing data systems to change by evolution not revolution (Tim Berners-Lee)
- An extension of the current Web, providing infrastructure for the integration of data.

Office of
PUBLIC SECTOR INFORMATION

research objectives

AKTive PSI

- Explore the potential for reuse of public sector information to stimulate the broader UK knowledge economy
- What is possible with public sector information using recent advances in web technology
- Research project without immediate commercial or service delivery drivers
- Build up a detailed picture of life in two London Boroughs, using as broad a collection of Public Sector Information as possible

Office of
PUBLIC SECTOR INFORMATION

initial observations

AKTive PSI

- Common evolution to bringing information online
- Initial phase: Scan historical & publish, capture new information as single entity (pdf, doc) & publish
- Current phase: Review information value chain, capture important information chunks in DB to help organise & publish for immediate needs.
- Re-use phase: Need to annotate information & publish such that others can use it in ways not imagined by the originator (using standards).

Office of PUBLIC SECTOR INFORMATION

ease of re-use

AKTive PSI

- Re-use inhibited by the internal information management issues
- Too many databases, separate schemas, minimal metadata, workflow for capturing metadata not established
- Seems to be no benefit to publish in a form for reuse
  - if it's core business there's no funding for re-use publishing
  - if it's not core (and licensable), then it becomes less important because the organisation focus is not on generating revenue.
- Are there simple initiatives that can aid ability to publish for re-use?

Office of
PUBLIC SECTOR INFORMATION

we aimed to show

AKTive PSI

- Show *how* information in existing databases can be made available in scalable **semantic knowledge bases**
  - Using latest semantic web languages to represent and query the data
- Show *how* all this data can be **linked** to create an extended knowledge network
- Show *how* **ontologies** can represent the given data
- Demonstrate examples of **added value**
- Investigate the suitability of **IPSV** for representing government data
- Identify **knowledge gaps** between existing databases, and how such gaps can be filled

Office of
PUBLIC SECTOR INFORMATION

some conclusions

AKTive PSI

- Small ontologies can do the job
  - Ontologies to limited domains
  - Can be integrated in various ways
- Use of ontologies
  - Data mapping and integration made easier
  - Helped to understand the data models
  - Flexibility of representation
  - Overall, we created around 19 million RDF statements
- Much can be gained when the data is integrated
  - Data about the same place or object is distributed across several databases and organisations
  - Data enrichment, consistency checks, better analysis
  - Better to integrate data from various sources, rather to duplicate it!
- Data access can be made easier
  - Mashups can be generated relatively easily
  - Search and retrieval across databases
  - Data can be published in “machine understandable” formats

Office of
PUBLIC SECTOR INFORMATION

developing the **re-use agenda**

- Giving things URI's that de-reference (e.g. clauses in copyright statements, licences, provisions in legislation)
- Making information available in a highly re-usable way (e.g. publishing data from the London, Belfast and Edinburgh Gazettes in RDF)
- Expressing rights and licensing information in a machine readable form
- Enabling machines to obtain permission to re-use (e.g. Click-Use web services)
- Developing a deeper understanding of the potential of the technology (e.g. a publicly available research space)

Office of
PUBLIC SECTOR INFORMATION

unlocking the potential of public sector information